

Final Report
Evaluating Nebraska's District Assessment Portfolios and
Recommending Model Assessments for Reading, Speaking, & Listening: 2002 - 2003

Prepared by

Chad W. Buckendahl, Ph.D.

James C. Impara, Ph.D.

Barbara S. Plake, Ph.D.

Abdullah Ferdous, M.A.

M. Kelly Haack, M.A.

Buros Institute for Assessment Consultation and Outreach

A Division of the
Oscar and Luella Buros Center for Testing

University of Nebraska - Lincoln

November 2003

Final Report
Evaluating Nebraska's District Assessment Portfolios and
Recommending Model Assessments for Reading, Speaking, & Listening: 2002 - 2003

Organization of this report

This report is separated into three sections. The first section describes briefly the evaluation and review process by which the district assessment portfolios were reviewed. This section describes the role of the National Advisory Committee for Assessment (NACA), the criteria for evaluating district assessment portfolios, the selection and training of the District Assessment Evaluation Team (DAET), and rescoring methods that were employed. The second section describes the results of the process including the number of assessment portfolios, the summary of the ratings, and the results of the rescoring activities. The third section contains conclusions and recommendations for next year's assessment portfolio review for Mathematics. Appendices to the report include the names and qualifications of the individuals who served on the District Assessment Evaluation Team (DAET) and the National Advisory Committee for Assessment (NACA).

Section 1: The evaluation and review process

The legislation that permits Nebraska school districts to design and use their own unique assessment systems for determining the achievement levels of their students on the Nebraska content standards in Reading, Speaking, and Listening has certain requirements. One requirement is that each district's assessment system has to be evaluated in terms of the technical quality of the assessments. A second requirement of the legislation requires that Nebraska's Department of Education (NDE) identify four assessment models that districts may adopt or adapt in designing future assessments.

NDE employed the Buros Institute for Assessment Consultation and Outreach (BIACO) to assist them in meeting both of these requirements. Thus, BIACO proposed a strategy for evaluating district assessment portfolios (technical manuals that describe a district's assessment system) and for the identification of the four assessment models. An early decision that related to both of these activities was that the quality of district assessments would be evaluated on the extent that six technical quality criteria were met. These technical quality criteria are: 1) Assessments match the standards; 2) Students have an opportunity to learn assessment content; 3) Assessments are unbiased and sensitive to cultural differences; 4) Assessments are at an appropriate level; 5) Assessment scores or decisions are reliable; and 6) Mastery levels are set appropriately. Each criterion was evaluated for each portfolio as being Met; Met - With Comment; Met - Needs Improvement; or Not Met.

NDE disseminated a scoring rubric to school districts in advance of the portfolio evaluation. This rubric characterizes the expectations for each rating level. Based on the combination of individual criterion ratings, a district receives an overall rating for the technical quality of their assessment portfolio. This overall rating places districts into one of five possible categories. These five categories are Exemplary; Very Good; Good; Acceptable, but Needs Improvement; and Unacceptable. The substance of the rubric was disseminated to districts in April 2002 in the form of the directions for completing their assessment portfolio. The rubric was also included in NDE's March, 2003 document, "Assessment Portfolio Instructions and Suggestions." The document was also available on the NDE web site (www.nde.state.ne.us). The requirements associated with how districts would be classified (e.g., Exemplary, Very Good,

etc.) was provided to districts in summer, 2001 and remains unchanged. The overall rating matrix was also provided in the NDE document referenced above.

BIACO and NDE collaboratively decided that the legislative requirement for four model assessments could be interpreted as four models for each of the six technical quality criteria. Identifying model assessment practices occurred during the first two years of the review process generating 24 models each year. The BIACO evaluations from the 2000-2001 and 2001-2002 portfolios resulted in 48 models that districts could adopt or adapt to their local systems. That is, there are eight models for each of the six technical quality criteria. Because it was unlikely that new models would be identified during this second round of review for the reading, speaking, and listening portfolios, additional review processes were added to provide greater evidence of validity and reliability.

Thus, the second principal outcome of the project was modified beginning in the 2002-2003 review to include greater numbers of portfolios that would be reviewed by multiple raters. Two extensions of the review process were added to meet this goal. The first extension of the review process included an automatic appeal process whereby all portfolios that received an overall rating of Unacceptable would be reviewed by the members of the National Advisory Committee for Assessment (NACA). The second extension of the review process included selecting a stratified (by reviewer) sample of 40 portfolios to independently double score. Two separate groups were employed to accomplish the goals of these reviews. The first group was the NACA. The second group was the District Assessment Evaluation Team (DAET).

NACA

The NACA had two principal functions. The first was to provide general advice on the activities associated with the evaluation of district assessment portfolios. Second, they had to conduct the automatic appeals process for districts that received overall ratings of Unacceptable.

In the previous two years, the NACA consisted of four members. One of these members was asked to resign from the NACA in June after he took a position with a commercially available test publisher. Because the nature of his employment represented a conflict of interest, it would not have been appropriate for him to continue to serve on the committee. With the short notice, we were unable to replace the individual in time for this year's training and automatic appeals process. However, a new member will be recruited for next year. Thus, the NACA for the 2002-2003 reviews consisted of three members. The three members were selected based on two criteria. The first criterion was that the individual be recognized nationally as having a high level of expertise in assessment, with a focus on state or local assessment programs. The second criterion was that the individual be one whose focus is on practical solutions to operational problems. These individuals were also required to be from outside Nebraska. This latter requirement reduced possible conflicts of interest when they selected the 48 models during the first two years or conducted automatic appeal reviews during the current year. Voting members of the NACA include a local school district assessment/research director, and two university faculty members who have published on classroom assessment and have worked in school systems. All members of the NACA are listed in Appendix A.

The NACA met on two occasions. The first meeting was during the DAET training session June 25-28, 2003. At this meeting, members were reminded about the characteristics of the Nebraska Assessment and Accountability System, various elements of the system addressed by BIACO or NDE were described, and the nature and variability of the assessment literacy of Nebraska school districts was discussed. The NACA members assisted in the training activities to articulate the criteria for identifying assessment strategies as models. The NACA also

participated in the training activities so that they would understand the evaluation rubric and had an opportunity to evaluate sample district portfolios to better understand the rating process as a whole. This prepared the NACA members for the later meeting where they would be conducting reviews during the automatic appeals process. The second meeting of the NACA occurred September 8-10, 2003 in Lincoln to conduct the automatic appeals of districts that received overall ratings of Unacceptable.

DAET

A team of 18 evaluators was recruited to apply the final technical quality rubric (NDE, 2003) to district assessment portfolios. This group of 18 people was named the District Assessment Evaluation Team (DAET). These individuals were recruited nationally. A primary selection criterion for membership on the DAET was that the individual had a strong background in assessment (preferably a Ph.D. in assessment or a related area such as school psychology). In addition, some knowledge or experience in Nebraska was desired, but not essential. All DAET members had their doctorate and a strong background in assessment. Many were either from Nebraska or were familiar with Nebraska school districts (ten were UNL graduates and two were recently on the UNL faculty). Some had assisted districts in developing their local assessment systems. All those who assisted districts were required to evaluate assessment portfolios of districts they had not assisted. Conflicts of interest were, to the extent possible, eliminated. The names and qualifications of the DAET are shown in Appendix B.

The DAET met June 25-28, 2003 to be trained in evaluating District Assessment Portfolios. Examples of performance for each criterion (Met, Met-Needs Improvement, and Not Met) from Reading portfolios from the 2000-2001 portfolios were used to calibrate the DAET to the evaluation rubric. Three districts' portfolios submitted to NDE in advance of the training were identified and used as part of the training. The training occurred in Lincoln over three and a half days.

The training process included an orientation to the Nebraska assessment and accountability system and a discussion of the six technical quality criteria. An electronic evaluation form designed to be securely accessed with unique usernames and passwords on the World Wide Web was created for the DAET members to rate and provide feedback to districts for each criterion that was rated as Met, With Comments, Met – Needs Improvement, or Not Met. Each criterion was defined and discussed in an attempt to come to consensus about what the criterion meant. Once there was moderate consensus on the meaning of the elements of each criterion (qualifications of who did the process, description of the process, results of the process), examples of each rating level were provided to illustrate what a district's performance may look like on the criterion. This was done for each rating for each criterion.

In this exercise, the DAET members were assigned to groups of 3-4 people. Each group looked at the first quality criterion (match to standards) and, as a group completed the electronic evaluation form. The results of each group's reactions were shared and discussed. This process was followed for each of the six quality criteria. The objective of this process was to put the initial abstract agreements about the meaning and interpretation of the criterion into an operational context. The discussion following the evaluation of each criterion often resulted in changes in what constituted acceptable procedures, what procedures might represent a need for feedback to the district, and what sorts of feedback would be appropriate to districts when feedback was needed. After these discussions, the first of the three reading, speaking, and listening (RSL) district assessment portfolios was examined. Following the independent ratings on the first RSL portfolio, the group discussed the characteristics of each criterion and came to

consensus on the rating and feedback (if necessary) the district should receive to improve their assessment practices. After the first reading assessment portfolio was evaluated, a second was reviewed using the same process. The second and third evaluations further clarified the acceptable procedures associated with the criteria and the nature of appropriate feedback.

During the first week of July 2003, portfolios were delivered to BIACO. BIACO assigned and repackaged the Assessment Portfolios and sent them to the DAET. There were a total of 19¹ reviewers among both the DAET and Buros staff. Most reviewers reviewed between 15 and 20 Assessment Portfolios. An additional element of the review process for the 2002-2003 portfolios was the “Take Note” indicator. This indicator alerted districts that a process or method employed may not be acceptable in the future.

Section 2: Results of the process

This section reports the results of the evaluation of district assessment portfolios, the automatic appeals process, and the consistency check across portfolio evaluators.

Evaluation of District Assessment Portfolios

Two hundred seventy-seven (277) assessment portfolios were received and reviewed. These portfolios represented more than 475 school districts. A number of consortia and collaborations were submitted which is why the number of portfolios was less than the number of districts represented by portfolios. Although there were a number of school districts that participated in a consortium (e.g., through their Educational Service Unit), many submitted their assessment portfolios independently. Districts that participated in a consortium submission all received the same rating for criteria 1 (match to standards), 3 (freedom from bias), 4 (developmental appropriateness), 5 (consistency in scoring), and 6 (appropriate mastery levels). For Criterion 2 (opportunity to learn), though, each district in a consortium submitted information about how they aligned their assessments with their local curriculum and instruction. This means that a district that participated in a consortium could receive a different rating on Criterion 2 and the overall rating based on their performance on this criterion than other districts that also participated in the consortium. In some instances, district indicated that they had participated in a consortium, but clearly used different processes and assessments. In these cases, districts were rated separately based on the portfolio information they supplied. Districts that did not have students at either the 4th or 8th grade do not receive a rating for the grade(s) that they do not have students.

As noted above, Assessment Portfolios for each grade level submitted were rated as being Exemplary, Very Good, Good, Acceptable, but Needs Improvement, or Unacceptable. An Assessment Portfolio that contained descriptions of the procedures used for three grades (typically Grades 4, 8 and 11) received a separate rating for each grade. In many cases, the same procedures were used at each grade level. In some cases, two grades (typically grades 4 and 8) used the same procedures, but the other grade level's (typically high school) procedures differed. In such cases, the ratings might differ on some criteria, potentially resulting in different overall classifications for two grade levels within the same district. In a limited number of cases, the procedures in all three grades differed, sometimes resulting in different classifications across all three grades. Some districts did not submit Assessment Portfolios for all three grades (e.g., Class 1 districts that have only elementary grades).

¹ Two DAET reviewers were identified in advance and were only used for the independent rescoring that occurred after the initial reviews by the other reviewers.

An Exemplary rating was given to each Assessment Portfolio that received ratings of Met on all six quality criteria. Met was interpreted as being Met; Met – With Comments; or Met – Needs Improvement with one exception. For Criterion 5, there was a distinction between Met or Met – With Comments, and Met – Needs Improvement. Because reliability is a necessary element for validity of inferences made about student performance, districts with a rating of less than Met – With Comments on this criterion received a lower overall classification. An overall rating of Very Good was given to districts that received a rating of Met on criteria one through four and a rating of at least Met – Needs Improvement on either Criterion 5 or Criterion 6. A rating of Good was actually more difficult to obtain because of the nature of the matrix that translates individual criterion performance into the overall rating. Most Assessment Portfolios were able to meet criteria one through four and they received at least a Met – Needs Improvement on either Criterion 5 or 6 to be classified as Very Good. Because some criteria are more critical than others, meeting any four criteria may not be sufficient to receive a rating higher than Acceptable, but Needs Improvement. Specifically, if a district met Criteria 1-4, but did not meet either Criterion 5 or 6, the overall rating would be Acceptable, but Needs Improvement. Many districts fell into this category because they did meet each of the first four criteria, but failed to meet both Criteria 5 and 6 (these are the two more technical criteria associated with consistency of scoring and setting mastery levels).

Table 1 shows the number of districts classified in each of the five overall rating categories. This number is not the same across all grade levels. This is because some districts received different ratings for their Assessment Portfolios for grades 4, 8, and 11. In other cases, a district may not have submitted portfolios for all three grades (e.g., Class I districts have only elementary grades).

Approximately 38%, 35%, and 41% of districts received ratings of Very Good at grade 4, 8, or 11 respectively. The percentage of districts that received overall ratings of Exemplary increased to 48%, 50%, and 51% for these same grade levels. Combined, approximately 87% of the districts received ratings of Very Good or Exemplary. Of the districts that received ratings of Unacceptable, very few received this rating because they submitted little or no documentation of the procedures and results associated with their local assessments. Some districts at grades 4 and 8 did not have any students at the grade level and should not be classified at any level. And other districts received an overall rating of Unacceptable because their documentation of Criteria 1 and/or 2 was evaluated by the reviewer as missing at least one important piece. The detailed ratings are shown in Table 1.

Table 1. Distribution of ratings of District Assessment Portfolios in Reading across all three grades.

Classification	4 th Grade	8 th Grade	11 th Grade
Exemplary	226	218	134
Very Good	179	153	106
Good	5	4	1
Acceptable, Needs Improvement	9	7	8
Unacceptable	48	51	11

As noted above, a number of assessment portfolios were received from consortia and all districts named as participants in a consortium were given the same rating on criteria 1, 3, 4, 5, and 6 because it was assumed (as per the instructions) that all districts in a consortium had used the same procedures. However, for Criterion 2 (Opportunity to Learn) each district was to submit separate information about how they addressed the criterion for ensuring that instruction in the content occurred prior to assessment. In some instances, this resulted in different ratings for districts that submitted in the same consortium.

Automatic Appeals/Double Scoring

One rater initially reviewed each Assessment Portfolio. Following an analysis of the initial ratings, the district portfolios that received an overall rating of Unacceptable were identified and organized for a second review by the National Advisory Committee for Assessment (NACA). This second review occurred September 10-12, 2003 in Lincoln. Members of the NACA met to calibrate on the evaluation rubric and conduct a second review of the portfolios that received overall ratings of Unacceptable. The NACA reviewed 140 portfolios during the automatic appeals meeting. Of these automatic reviews, 22 of the 140 (16%) were changed based on the second review. This means that 118 of 140 (84%) overall ratings were unchanged as a result of the second review. The NACA only reviewed criteria that received a Not Met during this review. This means that each portfolio identified for automatic appeal received a Not Met on either Criterion 1 or 2 (or both) and potentially other criteria that also received a Not Met within the portfolio. The majority of the portfolios that were identified for an automatic appeal did not meet Criterion 2 (opportunity to learn). Part of this criterion required districts to identify when the assessments were administered relative to instruction to provide evidence that students has been instructed in the content measured on the assessment before the assessment was administered. The issue of the timing of the assessment was not clearly articulated in the rubric that was disseminated to districts. It would have been unfair to hold the districts to a higher standard than they could reasonably assume based on the materials they received from NDE. Thus, following the automatic appeals reviews, districts that received a Not Met on criterion 2 only because there was no evidence of when the assessments occurred relative to instruction were given a Met – Needs Improvement along with a “Take Note” indication to alert the district that omitting this information in the future would not result in a similar rating.

In addition to the 140 portfolios that were reviewed as part of the automatic appeals process, a number of portfolios (n=40) were also double scored externally to verify the reliability and accuracy of the ratings. Portfolios reviewed in this part of the process originally received an overall rating higher than Unacceptable and would not have been reviewed as part of the automatic appeals process. For this part of the process, 2-3 portfolios were randomly selected from each DAET reviewer and rated without knowing the original overall rating by a second reviewer trained on the scoring rubric. This part of the review process occurred September 18-19, 2003 in Lincoln. The reviewers for this activity included two experienced DAET reviewers who were held out of the original rating team and two experienced DAET reviewers who part of the original rating team. To avoid a conflict of interest, the two reviewers that were part of the original rating team were not assigned the districts that they previously reviewed. The results of this review were that 25 of 40 (63%) resulted in the same rating (exact agreement) between the original rater and the second rater. Of the 37% that were not in agreement (n=15), the second review resulted in a lower rating in 12 (80%) of these cases suggesting the second reviewers may have been more stringent than the original reviewers. As this part of the process was intended only to monitor reliability, the original reviews were retained.

An additional 44 portfolios (16% of the total number of portfolios) were fully reviewed a second time by Bueros staff as an internal review to ensure that the scoring rubric was applied correctly. These portfolios were selected to ensure that the DAET members were applying the scoring rubric consistently across all districts in large collaborations. One reviewer was identified through the double scoring activity as systematically misapplying the scoring rubric. Most of this reviewer's ratings were incorrect, contributing greatly to the lower agreement for these second reviews. This part of the process was conducted with the second reviewers independently making a judgment about the rating of a criterion and then comparing their judgments to the original rating. When there was disagreement, the second reviewer's ratings were used. By combining the number of automatic appeal reviews with the additional double scoring activities, 224 of 277 portfolios (81%) were reviewed by at least two reviewers. A summary of these three sets of double scoring activities are summarized in Table 2.

Table 2. Results of overall rating double scoring activities for 2002-03 portfolio reviews.

<u>Process</u>	<u># agree (exact)</u>	<u># disagree</u>	<u>% agree (exact)</u>
Automatic Appeals	118	22	84%
Double Scoring (external)	25	15	63%
Double Scoring (internal)	30	14	68%
Totals	173	51	77%

District initiated Appeals/Resubmissions

Following the dissemination of assessment quality ratings, districts had the opportunity to appeal their rating on any criterion if they believed the reviewer made an error in the initial rating. Three portfolios representing eight districts submitted appeals. All three appeals were upheld and the districts' ratings on that criterion changed. Seven of the eight districts' overall ratings improved. It should be noted that none of the appeals were included in the automatic appeals process described earlier.

Districts were also allowed to resubmit information that was inadvertently omitted from the original portfolio. Seventy-two districts submitted additional information for consideration in their district's assessment quality rating. In most instances, the district's rating for a criterion or criteria changed as a result of the additional information. Note that automatic appeals, external and internal double scoring occurred prior to districts resubmitting information, so the level of agreement is between the original portfolio rating and a second review of the original portfolio without resubmitted information. Resubmissions will not be accepted in future years.

Evaluations

DAET members also responded to an evaluation about the process. The evaluation occurred during the training activities June 25-28, 2003. In summary, the DAET members who responded (16 of 16) said the training should include more anchor (example) portfolios to reflect the different levels of quality for each criterion and additional opportunities to rate assessment portfolios as a group to build consensus about how to handle specific rating questions. For greater functionality of the online rating form, reviewers requested to be able to go directly to a district's individual criterion, specifically criterion 2 within consortiums, as opposed to progressing sequentially through all six criteria. When asked how much time it took to complete each portfolio review the average time was more than 3 hours. This was higher than in previous years and not surprising. First, almost half of the districts in the state were involved in a

consortium or collaboration. This means that there were generally more materials to evaluate. For example, one portfolio filled four boxes and took over 20 hours to review. The second reason is that districts have grown in their knowledge of assessment over the past three years and are better prepared to present information to evaluate. Third, reviewers provided more detailed feedback to districts about appropriate methods and strategies given the district's assessment system and population characteristics. Compared to the first year of reading portfolios, very few portfolios could be reviewed very quickly (those with little or no information).

In summary, most of Nebraska's school districts submitted an assessment portfolio for review. Over 75% of the assessment portfolios received overall ratings of Exemplary or Very Good. As evidence from the number of districts that received Very Good as compared to those that received Acceptable, but needs improvement, there has been some progress on criteria 5 and 6. Districts tend to be able to earn at least a met-needs improvement rating on either criterion 5 or criterion 6. If criteria 1-4 are at least met-needs improvement, receiving at least a met-needs improvement on either of these last two criteria is enough to receive an overall rating of Very Good.

Section 3: Conclusions and Recommendations

The performance of districts in the third year of the portfolio review process suggests that districts have reached a point where expectations of the portfolios can be brought more in line with acceptable measurement practice. When the process began three years ago, the expectations for the portfolios were set at a level that would not be burdensome for the districts, but did not yet represent acceptable measurement practice. In modeling the performance language that districts are very familiar, they appear to be ready to move from the Progressing level to the Proficient level. This means that the low expectations need to also move up to be more in line with minimally acceptable measurement practice.

We recommend the following changes for the 2003-04 reviews when the assessment portfolios for mathematics will be re-submitted and evaluated.

1. Decisions about the rating rubric and overall rating matrix for evaluating districts should be finalized and disseminated by January, 2004. This will provide the districts with a time frame within which they can assemble their materials and organize them for the review process.
2. When districts submit their materials as a consortium, there is an expectation that each district in the consortium used the same procedures for all relevant criteria (i.e., all criteria except Criterion 2 – Opportunity to Learn). If this is not the case, each district should submit their assessment portfolio independently. There were some instances where consortiums were submitted where it was clear that districts in the consortium had not employed the same strategies across criteria 1, 3, 4, 5, and 6. Also, for districts that submit with a consortium, it is very important to submit separate information for criterion 2. In some situations, consortiums submitted only one piece of information for criterion 2 suggesting that all districts have the same curriculum and sequence instruction identically. The requirement that each district needs to individually provide evidence for criterion 2 should be emphasized in the portfolio instructions.

3. NDE materials also need to be revised to clarify the distinction between collaborative and consortia submissions. There was a great deal of confusion among districts in this distinction this year. Many districts that worked with other districts using similar development processes, but did not use the same assessments, or have the same results identified themselves as a consortium. These needed to be disentangled in order to provide separate ratings for each portfolio. The reverse situation also occurred where districts had the same materials, processes, assessments, and results copied in 10-15 portfolios that were collaboratively submitted rather than as a consortium.
4. A large number of districts are participating in consortiums. For small districts this is especially encouraging as assessment literacy capacity continues to evolve collaboratively. Although the efficiency of these systems is an advantage, it is appropriate to caution the districts in these consortia about the limitations of centralized systems. There are some concerns that assessment literacy development may become stagnant if the responsibility for the local assessment system is too far removed from the local system. Additional professional development in effectively using information locally will encourage integration of curriculum, instruction, and assessment. It will also help to address the anecdotal concerns that some districts expressed in their portfolios that in their perception, local assessment processes were disconnected from teaching and learning and are viewed as an external layer.
5. Districts have provided examples of their local assessments during the past three years. Although the portfolio evaluations over the past three years have rated the process and outcomes of those processes with respect to technical quality, the actual items or tasks used to measure student performance have not been reviewed at the state level. We recommend collecting assessments from districts and reviewing them for content and structure. This review could be conducted by a team of content experts and item writing experts to provide additional feedback to school districts on their local assessments.
6. Additional professional development opportunities are needed for districts to continue to grow in their assessment literacy skills. We recommend greater utilization of UNL's assessment cohort "graduates" for some of these activities. Cohort members have developed skills that they can share with their colleagues. This broadens the opportunities for professional development and adds greater credibility because most cohort members have extensive classroom experience that they would be able to integrate into training.
7. We also recommend greater utilization of the National Advisory Committee for Assessment. The individuals who serve on this advisory committee bring broad experiences from classroom assessment, large district assessment, state assessment, and higher education that would be beneficial to policymakers. Their role during the first few years has been narrowly defined in the context of the portfolio review process, but could be expanded to take advantage of the perspectives that they can offer to Nebraska's unique assessment and accountability system.
8. Although there were improvements in overall ratings, these results do not yet suggest that district portfolios that received Very Good or Exemplary represent minimally acceptable

measurement practice. The rubric and overall rating matrix designed in 2000 and updated in 2002 were intended to be very lenient during the initial development process because it was anticipated that there was a low level of assessment literacy capacity across districts. Districts have now had four years of experience and based on the districts' performance, it is appropriate to expect that the local assessment systems more closely align with acceptable measurement practice. The current rating rubric and overall rating matrix should be revised to represent these expectations. It is also important to bring these expectations more in line with acceptable measurement practice as these data are now being used as part of state and federal accountability systems. To be consistent with the formative nature of the state's system, we recommend that these updated expectations be disseminated by January, 2004 to allow districts to see the longer term goals of their local assessment systems. However, the state is encouraged to allow the districts adequate opportunities to revise their assessments and receive feedback on the greater expectations before they are held accountable for them.

Overall we believe the evaluation process was an improvement from the first two years. Districts have been very responsive to the technical quality criteria and continue to improve their practices locally. As Nebraska's Assessment and Accountability System continues to evolve, expectations for assessment quality will increase along with assessment literacy capacity that has been developed over the past few years.

Appendix A

Names and qualifications of National Advisory Committee for Assessment

Name	Current Position	Other qualifications
Dr. Jeri Benson	Professor, University of Georgia (GA)	Editor of major journal in applied measurement, expert in test validity.
Dr. Susan Brookhart	Assessment Consultant (MT)	Recently of Duquesne University (PA). Has major publications related to classroom assessment
Dr. Joe Wilhoft	Director, Planning, Assessment, and Evaluation, Tacoma Public Schools (WA)	Works with Washington State Assessment (Technical Advisory Committee, and other school districts).

Appendix B

Names and qualifications of District Assessment Evaluation Team

Name	Current Position	Other qualifications
Dr. Debbi Bandalos	Associate Professor, University of Georgia (GA)	Teaches measurement courses and has worked with Nebraska school districts on assessment development.
Dr. Laura Barnes	Associate Professor, Oklahoma State University (OK)	Teaches measurement courses, works with test publishers on norming projects.
Dr. Jennifer Fager	Director, Curriculum and Assessment, Central Michigan University (MI)	Works on a variety of assessment and evaluation projects with school districts and higher education.
Dr. Gerald Giraud	Assistant Professor, Nebraska Methodist University (NE)	Teaches measurement and statistics and has worked on assessment projects with Nebraska school districts.
Dr. H. Guy Glidden	Assessment Consultant (KS)	Retired director of assessment for Wichita (KS) Public Schools. Experience in developing and validating tests at the district level.
Dr. Jorge Gonzalez	Assistant Professor, Center for At-Risk Students, UNL (NE)	Worked with Buros Institute of Mental Measurements checking facts and writing descriptions of commercially available tests.
Dr. Jessica Jonson	University Wide Assessment Coordinator, UNL (NE)	Works with all UNL departments to assess student outcomes in a variety of programs.
Dr. Steve Lehman	Assistant Professor, Utah State University (UT)	Cognition and learning specialist with experience in evaluation, instrument design and validation.
Dr. Sherri Miller	Program Director, ACT (IA)	Works on numerous assessment projects with schools and other agencies.
Dr. Lori Nebelsick-Gullett	Assessment Consultant (CO)	Works with state and local assessment systems. Also teaches as part of the assessment cohort program at UNL.
Dr. John Nietfeld	Assistant Professor, North Carolina State University (NC)	Cognition and learning specialist with experience in instrument design and validation.

Dr. Lynn Nietfeld	Director, Athletic Department's Academic Counseling Center, North Carolina State University (NC)	Experience with assessment design and validation. Has worked with both school districts and higher education.
Dr. Lori Olafson	Assistant Professor, University of Nevada – Las Vegas (NV)	Reading assessment specialist with experience working with school districts in the United States and Canada.
Dr. Gregg Schraw	Professor, University of Nevada – Las Vegas (NV)	Cognition and learning specialist with expertise in instrument design and validation.
Dr. Howard Stoker	Assessment Consultant (TN)	Faculty member at Florida State for over 30 years, directed state testing program in FL, has worked with numerous state, and local, agencies on testing issues.
Dr. Elisabeth Sundermeier	Staff Counselor, Rhode Island Counseling Center (RI)	Worked with Buros Institute of Mental Measurements fact checking reviews for commercially available tests.
Dr. Ed W. Wiley	Assessment Consultant (CA)	Educational measurement and policy experience with California's state assessment and accountability system.